

# Crawling the cosmic network: identifying and quantifying filamentary structure

Nicholas A. Bond,<sup>\*</sup> Michael A. Strauss and Renyue Cen

*Princeton University Observatory, Department of Astrophysical Sciences, Princeton, NJ 08544, USA*

Accepted 2010 July 6. Received 2010 July 6; in original form 2010 March 5

## ABSTRACT

We present the Smoothed Hessian Major Axis Filament Finder (SHMAFF), an algorithm that uses the eigenvectors of the Hessian matrix of the smoothed galaxy distribution to identify individual filamentary structures. Filaments are traced along the Hessian eigenvector corresponding to the largest eigenvalue and are stopped when the axis orientation changes more rapidly than a preset threshold. In both  $N$ -body simulations and the Sloan Digital Sky Survey (SDSS) main galaxy redshift survey data, the resulting filament length distributions are approximately exponential. In the SDSS galaxy distribution, using smoothing lengths of 10 and 15  $h^{-1}$  Mpc, we find filament lengths per unit volume of  $1.9 \times 10^{-3}$  and  $7.6 \times 10^{-4} h^2 \text{ Mpc}^{-2}$ , respectively. The filament width distributions, which are much more sensitive to non-linear growth, are also consistent between the real and mock galaxy distributions using a standard cosmology. In SDSS, we find mean filament widths of 5.5 and 8.4  $h^{-1}$  Mpc on 10 and 15  $h^{-1}$  Mpc smoothing scales, with standard deviations of 1.1 and 1.4  $h^{-1}$  Mpc, respectively. Finally, the spatial distribution of filamentary structure in simulations is very similar between  $z = 3$  and  $z = 0$  on smoothing scales as large as 15  $h^{-1}$  Mpc, suggesting that the outline of filamentary structure is already in place at high redshift.

**Key words:** methods: data analysis – surveys – cosmology: observations – large-scale structure of Universe.

## 1 INTRODUCTION

Observational evidence for filamentary structures in the large-scale distribution of galaxies was first presented in galaxy redshifts surveys (e.g. Thompson & Gregory 1978; Davis et al. 1982; de Lapparent, Geller & Huchra 1986; Sathyaprakash et al. 1998; Colless et al. 2001; Gott et al. 2005). When similar structures were seen in cosmological  $N$ -body simulations of the dark matter distribution (e.g. Bond, Kofman & Pogosyan 1996; Sathyaprakash, Sahni & Shandarin 1996; Aragón-Calvo et al. 2007; Hahn et al. 2007a), a picture of a vast ‘cosmic web’, in which filaments skirted the boundaries of voids and were connected by galaxy clusters, began to emerge. These filaments are thought to provide pathways for matter to accrete on to galaxy clusters (e.g. Tanaka et al. 2007) and to torque dark matter haloes to align their spin axes (Hahn et al. 2007a,b, 2009). Filaments also produce deep potential wells and will give rise to a gravitational lensing signal on the largest scales (Dietrich et al. 2005; Massey et al. 2007). A number of authors have claimed detections of filaments using weak lensing (e.g. Kaiser et al. 1998; Dietrich et al. 2005; Massey et al. 2007), but simulations predict that structure along the line of sight should produce shear comparable to that of the target filaments (Dolag et al. 2006) and the

evidence remains far from conclusive. In addition, the formation of filaments is accompanied by gravitational heating, which gradually increases the temperature of the intergalactic medium over time and produces the so-called warm-hot intergalactic medium by  $z = 0$  (e.g. Cen & Ostriker 1999).

Perhaps the simplest and most effective means of identifying *clusters* in discretely sampled fields, such as redshift surveys and  $N$ -body simulations, is the friends-of-friends algorithm (FOF; Huchra & Geller 1982), in which particle groups are assembled based on the separation of nearest neighbours. These FOF structures can then be quantified with ‘Shapefinders’, statistics which measure the length, breadth and thickness of structures and are related to the Minkowski functionals (Sahni, Sathyaprakash & Shandarin 1998). Sheth et al. (2003) have developed an algorithm for computing the Shapefinders on structures at an arbitrary density threshold. Many of those found in data and simulations are indeed filamentary, but FOF algorithms are optimized for structures that lie above a set density threshold, a condition approximately met by clusters at the present epoch. Filaments and walls, however, are not bound and a strict density cut alone would not provide clean samples of such structures.

Another algorithm, called the Skeleton (Novikov, Colombi & Doré 2006; Sousbie et al. 2008a,b), identifies filaments by searching for saddle points in a density field and then following the density gradient along the filament until it reaches a local maximum.

<sup>\*</sup>E-mail: nbond@physics.rutgers.edu

Although it appears to be effective at making an outline of the cosmic network, it lacks an intuitive definition of filament ends. Aragón-Calvo et al. (2008) were also not able to provide such a definition, but has been successful at tracing the filament network in cosmological simulations using watershed segmentation (see also Platen, van de Weygaert & Jones 2007) and a Delaunay tessellation density estimator (Schaap & van de Weygaert 2000). If we wish to analyse filament length distributions or their spatial relationship to clusters, it is important to separate individual filaments in the cosmic web. Structure-finding techniques that only detect filaments between galaxy cluster pairs (e.g. Colberg, Krughoff & Connolly 2005; Pimbblet 2005; Gonzalez & Padilla 2010) would present a biased view of the filament–cluster relationship.

An early technique for identifying filaments in two-dimensional data was developed by Moody, Turner & Gott (1983) that works on a similar principle to the algorithm described in this paper. It divides the density field into a pixelized grid and identifies as filament elements any grid cell that has a larger density than its immediate neighbours along two of the four axes (including the two coordinate axes and two axes at  $45^\circ$  angles to the grid) through the grid cell. The algorithm was run on the Shane–Wirtanen galaxy count catalogue (Seldner et al. 1977) but has not been developed further. A later algorithm, presented by Dave et al. (1997), works on a similar principle, identifying ‘linked sequences’ using the eigenvectors of the inertia tensor. The authors found that the algorithm was poor at discriminating between cosmological models using CfA1-like mock galaxy catalogues, primarily because of the small number of galaxies in the catalogues.

In Paper I, we used the distribution of the Hessian eigenvalues of the smoothed density field ( $\lambda$ -space) on a grid to study three types of structure: clumps, filaments and walls. Filaments were found in the  $\lambda$ -space distributions at a variety of smoothing scales, ranging at least from 5 to  $15 h^{-1}$  Mpc, in both  $N$ -body simulations and the galaxy distribution measured by the Sloan Digital Sky Survey (SDSS; York et al. 2000). Furthermore, filaments were found to dominate the large-scale distribution of matter using smoothing scales of  $10$ – $15 h^{-1}$  Mpc, giving way to clumps with  $\sim 5 h^{-1}$  Mpc smoothing.

The fact that the eigenvalues of the Hessian can be used to discriminate different types of structure in a particle distribution is fundamental to a number of structure-finding algorithms (e.g. Colombi, Pogosyan & Souradeep 2000; Aragón-Calvo et al. 2007; Hahn et al. 2007a; Forero-Romero et al. 2009). However, the relationship between  $\lambda$ -space and a particular structure is not always trivial. For example, one might think that a filamentary grid cell would have two positive and one negative eigenvalues. This will be true near the centre of a filament connecting two overdense filament ends, but in the vicinity of the overdensities or in the case that the filament ends at an underdensity, all three eigenvalues will become negative. In addition, when working with a smoothed density field, these criteria select regions that are near clumps and do not necessarily lie along the filament. Finally, these criteria disregard the structure’s width – for example, the regions away from the centre of the filament may have positive values of  $\lambda_2$ .

In this paper, we will describe a procedure to identify filaments in the three-dimensional galaxy distribution using an algorithm called the Smoothed Hessian Major Axis Filament Finder (SHMAFF) and compare their properties in cosmological  $N$ -body simulations to those in the SDSS galaxy redshift survey. We describe our methodology, which uses the eigenvalues and eigenvectors of the smoothed Hessian matrix (see Bond, Strauss & Cen 2010, hereafter Paper I), in Section 2. In Section 3, we run the code with a range of possi-

ble input parameters and justify our choices for each. We discuss the behaviour of the algorithm when used on Gaussian random fields in Section 4, allowing us to distinguish those features of the large-scale distribution of matter that are a direct consequence of the non-linear growth of structure. In Section 5, we use mock galaxy catalogues to estimate the incompleteness and contamination rates of filament samples and then use these quantities to interpret the distribution of filaments found in the SDSS (Section 6). In Section 7, we summarize our results and discuss the implications of our findings.

## 2 FINDING INDIVIDUAL FILAMENTS

Filaments, clusters and walls all present sharp features in the density field along at least one of their principal axes. In Paper I, we described a procedure to generate a matrix of Gaussian-smoothed second derivatives of the density field (the Hessian matrix) at each grid cell, computing its eigenvalues,  $\lambda_i$  (defined such that  $\lambda_1 < \lambda_2 < \lambda_3$ ), and eigenvectors,  $A_i$ . For the testing and development of the algorithm, we ran a series of cosmological  $N$ -body simulations, using a particle-mesh code with  $\Omega_m = 0.29$ ,  $\Omega_\Lambda = 0.71$ ,  $\sigma_8 = 0.85$  and  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.69$  (see Paper I for details). The simulation is performed within a  $200 h^{-1}$  Mpc box with  $512^3$  particles, each with mass,  $m_p = 4.77 \times 10^9 h^{-1} M_\odot$ .

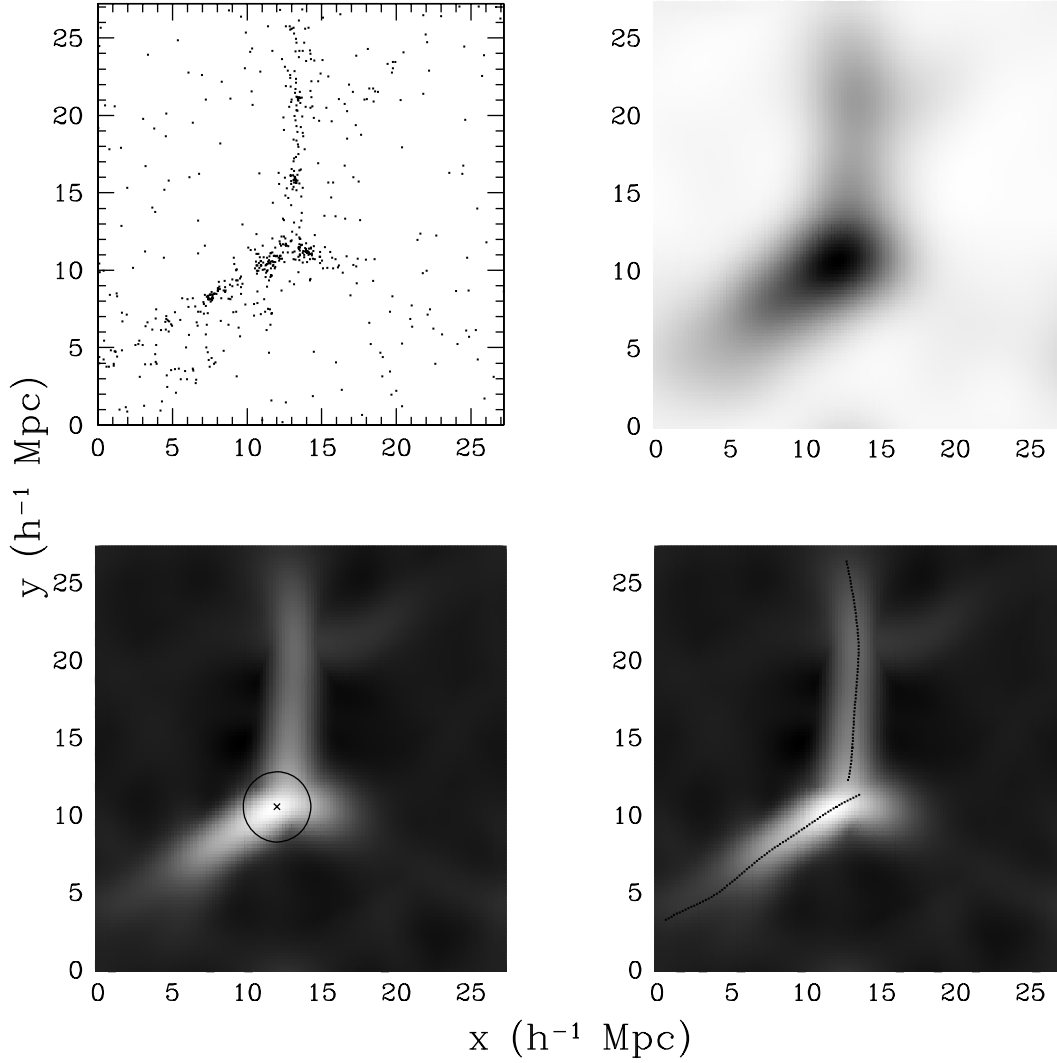
In order to generate a three-dimensional distribution of mock galaxies, we first identify dark matter haloes within the particle distribution using the HOP algorithm (Eisenstein & Hut 1998) and then populate them using the halo occupation distribution and parametrization of Zheng, Coil & Zehavi (2007, see Paper I for details). The resulting mock galaxy distribution is smoothed using a Gaussian kernel and its second derivatives, yielding a  $128 \times 128 \times 128$  grid with Hessian eigenvalues and eigenvectors in each cell. In Fig. 1, we plot a slice from the simulation  $10 h^{-1}$  Mpc deep and  $27.21 h^{-1}$  Mpc on a side, chosen to encompass a prominent filamentary structure. Shown are the galaxies (upper left), galaxy density map (upper right) and  $\lambda_1$  map (lower left and lower right), smoothed with a  $l = 2 h^{-1}$  Mpc kernel to bring out the filament. The structure appears most clearly in  $\lambda_1$ , so we construct a list of grid cells,  $G$ , ordered by an increasing value of  $\lambda_1$ . Before marking the first filament, we remove from  $G$  all grid cells that satisfy any of the following criteria:

$$\begin{aligned} \lambda_1 &> 0 \\ \lambda_2 &> 0 \\ \rho &< \bar{\rho}, \end{aligned} \tag{1}$$

where  $\bar{\rho}$  is the mean density of objects making up the density field. The  $\lambda_1$  and  $\lambda_2$  thresholds follow from the definition of a filament – the density field must be concave down along at least two of the principal axes.

The first element in  $G$  (the most negative in  $\lambda_1$ ) is marked with a cross in the lower left panel of Fig. 1. From this starting point, we trace out the filament in both directions of the ‘axis of structure’ (parallel and antiparallel to  $A_3$ ), taking steps equal to the grid scale of  $1.5625 h^{-1}$  Mpc. Subsequent filament elements are not constrained to lie on the grid, so we use a third-order polynomial interpolation scheme (Press, Flannery & Teukolsky 1986) on the grid to obtain the local Hessian parameters. If, at any point along the filament, the angular rate of change of the axis of structure exceeds a threshold,  $C$ , we stop tracing and mark the point as a filament end. The stopping condition at step  $m$  is given by

$$|A_{3,m} \times A_{3,m-1}| > \sin(C \Delta), \tag{2}$$



**Figure 1.** Slice from a cosmological simulation  $10 h^{-1}$  Mpc deep and  $27.21 h^{-1}$  Mpc on a side, encompassing a prominent filamentary structure. Shown are the galaxies (upper left), density map (upper right) and  $\lambda_1$  map (lower two panels), where smoothing is performed on a scale of  $l = 2 h^{-1}$  Mpc. The cross in the lower left panel indicates the minimum value of  $\lambda_1$  on the slice. This will be the starting point for the first filament traced by the algorithm. The circle around the cross has a radius equal to the removal width (see the text). The solid lines in the lower right panel indicate the filaments as traced by a 2D version of SHMAFF.

where  $\Delta$ , the grid cell size, is also the size of each step. The filament finder will also stop and mark a filament end if it passes into a cell that satisfies one or more of the criteria specified in equation (1). In the lower right panel of Fig. 1, we show the filaments that result from a sample run of SHMAFF on a  $27 \times 27 \times 10 h^3 \text{ Mpc}^{-3}$  slice from the dark matter particle distribution in our cosmological simulation.

For each step along a filament, all grid cells within a width,  $W$ , of the most recently chosen filament element, are removed from  $G$ , where

$$W_i = K \sqrt{\frac{-\rho_i}{\lambda_{1,i}}}. \quad (3)$$

In order to avoid tracing a filament more than once, subsequent filaments cannot start within one of the removed cells. They may, however, extend into a removed cell, so long as the cell is not excluded by any of the criteria given in equations (1) and (2). For a cylindrical filament with a Gaussian cross-section extending into a zero-density background, a value of  $K = 1$  should exclude those parts of the structure that are not already excluded by equation (1).

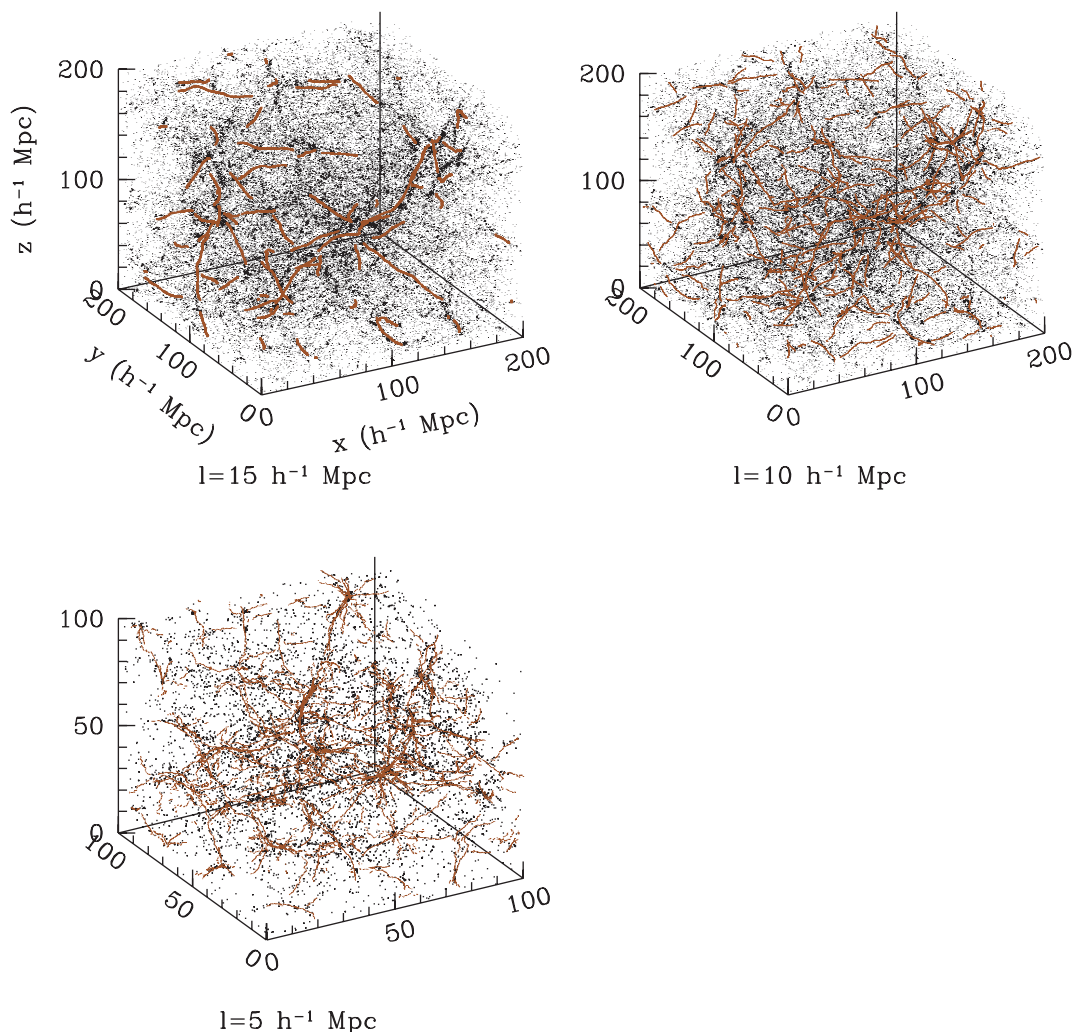
The filaments traced by the above algorithm may be offset from the ridges in the initial point field because of the finite resolution of the grid. Thus, we adjust the position of a filament element,  $j$ , based on the average *perpendicular displacement* of nearby grid cells from the filament axis,

$$\bar{\Delta} s_j = \frac{\sum_{i=1}^N \mathbf{R}_i}{N}, \quad (4)$$

where

$$\mathbf{R}_i = \hat{\mathbf{A}}_j \times (\hat{\mathbf{A}}_j \times (\mathbf{x}_j - \mathbf{x}_i)). \quad (5)$$

Here,  $\hat{\mathbf{A}}_j$  is the unit vector along the axis of structure (with an arbitrary sign) and  $N$  is the number of objects in the initial point field that are within a smoothing length. Application of the centring algorithm can result in fragmented filaments when shot noise is non-negligible, so we will not run it on point distributions with very sparse sampling, such as the SDSS galaxy distribution.



**Figure 2.** Filaments found in the  $z = 0$  dark matter distribution of a cosmological simulation are plotted (in red) over a subsample of dark matter particles (black). Filaments are found in a density field smoothed with the kernel length indicated below each box. For  $l = 5 h^{-1}$  Mpc, we only show an octant of the full simulation box, as the full filament distribution is so rich that the figure for the full box would be too crowded.

### 3 FILAMENT-FINDING PARAMETERS

In the filament-finding routine described in Section 2, there are two free parameters, the curvature criterion for identifying the filament ends,  $C$ , and the width of filament removal,  $K$  (see equation 3). In principle, the optimal values of these parameters can be functions of the smoothing scale, redshift or type of tracer (e.g. galaxies, dark matter particles), so it is important to understand their impact on the algorithm's output. In this section, we will test the performance of the code on the distribution of dark matter particles in our cosmological simulation as a function of  $K$ ,  $C$  and the sampling rate.

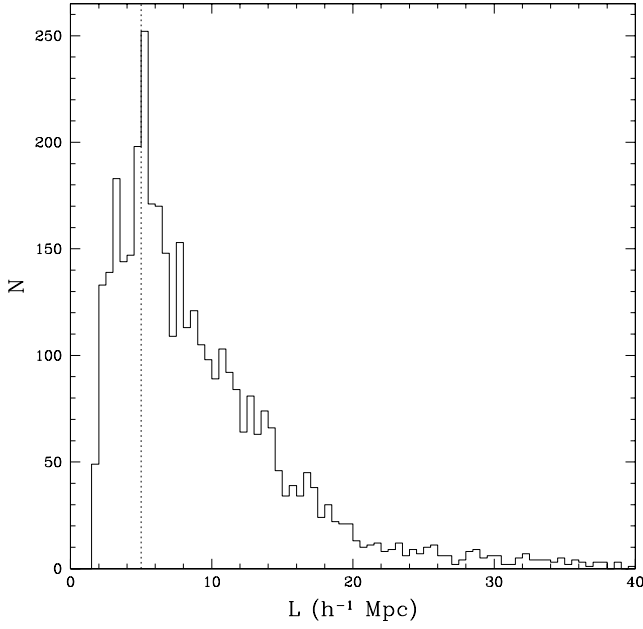
In Fig. 2, we show a run of the filament finder on the  $z = 0$  dark matter distribution of the cosmological simulation, using  $C = 40^\circ l^{-1}$  and  $K = 1$ . Output is shown for smoothing with  $l = 15, 10$  and  $5 h^{-1}$  Mpc, illustrating the scale dependence and coherence of the cosmic network. Any given filament will be found on a range of scales, depending on its width and length, but as the smoothing length is made smaller, the filament will be broken up into substructures which will themselves be filamentary or clump-like (see figs 15 and 16 in Paper I).

#### 3.1 Sampling the filament length distributions

Before defining the parameters that are used to find filaments, we must decide on what we are willing to accept as a real filament. An isolated spherical overdensity should not be viewed as a filament, but the filament finder would treat it as a very short ridge, tracing it from its centre until random fluctuations caused the axis of structure to deviate more than  $C$ , producing a ‘short filament’. Fig. 3 shows the raw distribution of filament lengths for our dark matter simulation shown in Fig. 2, using  $C = 30^\circ l^{-1}$ ,  $K = 1$  and a smoothing length of  $5 h^{-1}$  Mpc. Not surprisingly, the distribution exhibits a dramatic drop-off below a smoothing length. With this in mind, we hereafter discard filaments whose lengths are shorter than the smoothing length as non-physical.

#### 3.2 The $C$ parameter

The traditional picture of large-scale structure as a ‘cosmic web’ (Bond et al. 1996) suggests that filaments are connected, one-dimensional strands that end abruptly at their points of intersection. As one filament begins and another ends, the local axis of



**Figure 3.** Length distribution for a sample of filaments found in the  $z = 0$  smoothed dark matter distribution (smoothing kernel width of  $l = 5 h^{-1}$  Mpc). The dashed line indicates the smoothing length, below which filaments are removed from the sample.

structure should change the direction rapidly. The  $C$  parameter denotes the maximum angular rate of change in the axis of structure along a filament. If this threshold is exceeded, filament tracing is stopped.

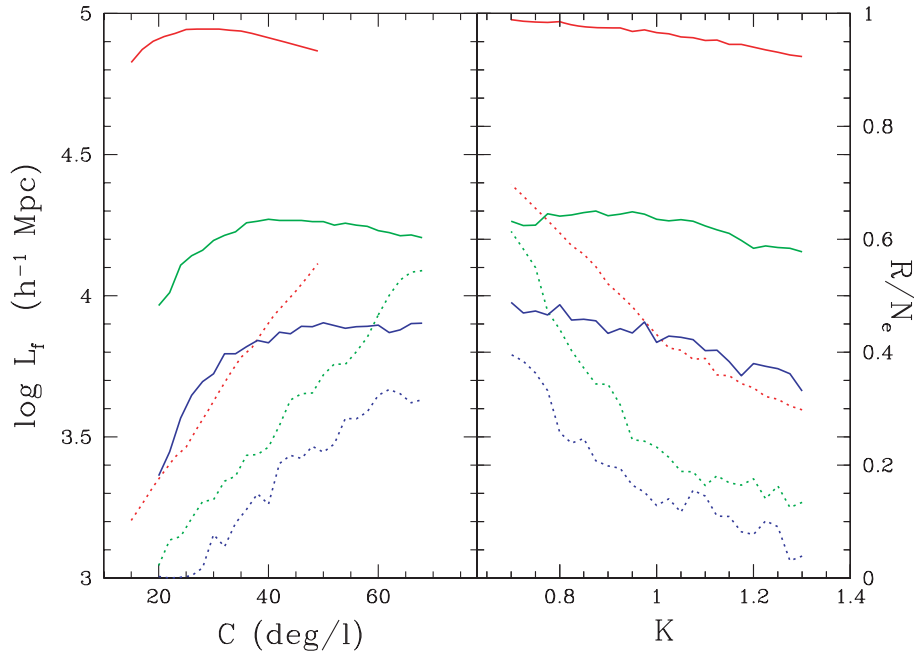
In order to test the sensitivity of the output filaments to the value of the  $C$  parameter, we set  $K = 1$  and generated filament networks in the  $N$ -body simulation with a range of  $C$ . In all of these tests,

increasing the value of  $C$  led to an increase in the average length of the filaments and a decrease in the total number of filaments found. If the curvature criterion is not strict enough, a filament will be traced past its vertex and into another filament. Since our algorithm only prevents filaments from starting within previously identified filaments (they are allowed to cross one another), this can lead to double detections of filaments. We can obtain a rough count of these double detections by comparing filament elements to one another, where a filament element is defined as a single step (of interval,  $\Delta$ ) on the grid. In other words, for each step along a given filament, we find the closest filament element that is not a member of that same filament. If the closest filament element is within a smoothing length and has an axis of structure within  $C$ , then the original element is labelled a ‘repeat detection’. The total number of repeat detections in an output filament network is denoted by  $R$ . The total length of the network at this scale is therefore given by

$$L_f = (N_e - R)\Delta, \quad (6)$$

where  $N_e$  is the total number of filament elements found and  $\Delta$  is the step size taken by the filament finder. Non-filamentary regions of space have already been excluded by the criteria in equation (1), so an optimum set of parameters will maximize  $L_f$  while minimizing  $R$ .

In the left-hand panel of Fig. 4, we plot both the fraction of repeat detections ( $R/N_e$ , dotted lines) and the total length of the network ( $L_f$ , solid lines) as a function of  $C$ . On all smoothing scales, the fraction of false positives increases steadily with increasing  $C$ , with no obvious breaks or minima. The total length, however, tends to rise until it reaches a maximum, after which point it either flattens or falls slowly. This suggests that, as long as the curvature criterion is above a critical value, the algorithm will trace out the entire filament network. Since the fraction of false positives increases with  $C$ , we will hereafter use a curvature criterion near this value; that is,  $C = 50, 40$  and  $30^\circ l^{-1}$  for  $l = 15, 10$  and  $5 h^{-1}$  Mpc, respectively.



**Figure 4.** Total length of the output filament network (solid lines) and the fraction of ‘repeat detections’ (dotted lines) as a function of the curvature criterion,  $C$  (left) and  $K$  (right). The value of  $C$  determines the filament ending points and the value of  $K$  determines the removal width (see equation 3). Filaments were identified on three different smoothing scales,  $l = 15 h^{-1}$  Mpc (blue),  $l = 10 h^{-1}$  Mpc (green) and  $l = 5 h^{-1}$  Mpc (red). The total length of the network (after removing repeat detections) maximizes at a value of  $C$  that depends on the smoothing length.

### 3.3 The $K$ parameter

As each filament is found, we wish to remove from the grid as much of it as possible without preventing the detection of further real filaments. Using the previously determined critical values of  $C$ , we ran the filament finder with a range of  $K$  and computed the total length of the filament network and the fraction of repeat detections as a function of  $K$ . The results are shown in the right-hand panel of Fig. 4. All of the curves are monotonic, with repeat detections and the network length decreasing with increasing  $K$ . Hereafter, we will set  $K = 1$  because it yields  $R/N_e \lesssim 20$  per cent.

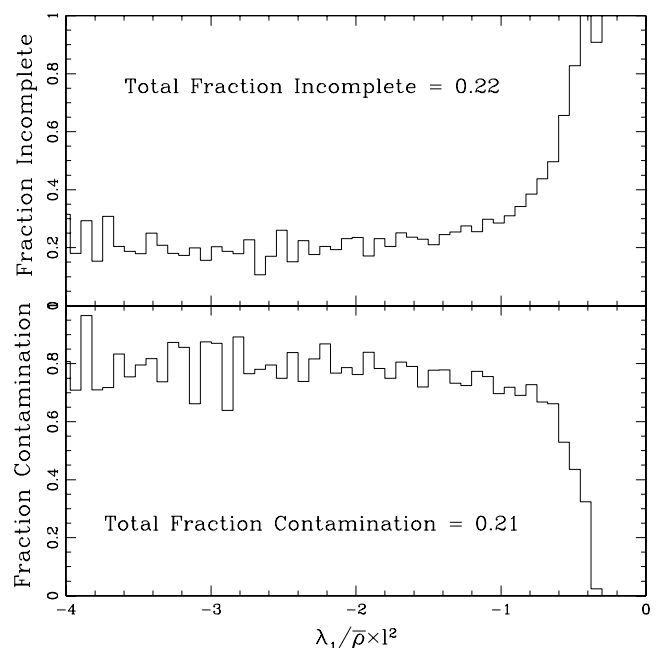
### 3.4 Effects of sparse sampling

In real galaxy catalogues, the number of galaxies per smoothing volume will sometimes be small and it is important to understand the impact of shot noise on the algorithm's ability to trace the filament network. In a density field with sparse sampling, shot noise will create spurious filament detections in addition to the 'repeat detections' described in Section 3.2. We have an effectively shot-noise-free density field in the dark matter particle distribution (with the simulation using a mean dark matter particle density of  $17 \text{ particles } h^3 \text{ Mpc}^{-3}$ ), so we perform sparse sampling on this field and use the complete particle distribution as a standard for comparison. We construct three such data sets, sampled to densities of  $5 \times 10^{-3}$ ,  $2 \times 10^{-3}$  and  $1 \times 10^{-3} \text{ particles } h^3 \text{ Mpc}^{-3}$ , respectively, matching the densities of the real galaxy samples to be presented in Section 6. For each sample, we recompute the SHMAFF parameters and run the filament finder on all three smoothing scales, using the parameters derived in previous sections. We will call a 'false positive' any filament element found in the sparsely sampled data whose nearest neighbouring element in the 'true' filament network is more than a smoothing length away or does not have an axis of structure within an angle equal to  $C \times l$ . Similarly, incompleteness is quantified by counting the filament elements in the 'true' network that have no counterparts in the sparse-sampled one.

As illustrated in Fig. 5, the incompleteness and contamination rates of individual filament elements are strong functions of  $\lambda_1$  for the 'weakest' edges, but these make up only a small fraction of the filament network. Our tests (not shown) suggest that the incompleteness and contamination rates are  $\lesssim 20$  per cent so long as there are an average of  $\gtrsim 5$  particles within spheres of radius equal to the Gaussian smoothing length. See Section 5 for a more detailed analysis of completeness and contamination in mock galaxy samples.

## 4 FILAMENTS AS NON-GAUSSIANITIES

Gaussian random fields serve as an important reference point if we wish to distinguish the consequences of the non-linear growth of structure from phenomena seen only in the linear regime. We know that Gaussian random fields are not filamentary and one might question why we should find *any* filaments in such a distribution. Bear in mind, however, that the SHMAFF algorithm traces any negatively curved region and these conditions will certainly be met by some of the overdensities in a Gaussian random field. We demonstrated in Paper I that although the smoothed  $\lambda_1$  distributions appeared 'filamentary' in both a Gaussian random field and the evolved dark matter distribution, the latter showed alignment between the axis of structure and these filamentary minima in  $\lambda_1$ . The filament-finding algorithm enables us to follow the axis



**Figure 5.** Incompleteness and contamination of filament pixels as a function of  $\lambda_1$  (scaled to the mean density and smoothing length) in the filament elements of a dark matter particle distribution sparse-sampled to match the density of  $M_r < -20$  galaxies ( $5.0 \times 10^{-3} h^3 \text{ Mpc}^{-3}$ ). The sparse-sampled field was smoothed on a scale of  $5 h^{-1} \text{ Mpc}$ , yielding 2.6 particles per smoothing volume, and its filaments were compared with those in the full dark matter particle distribution smoothed on the same scale. Their total incompleteness and contamination rates were both  $\sim 20$  per cent.

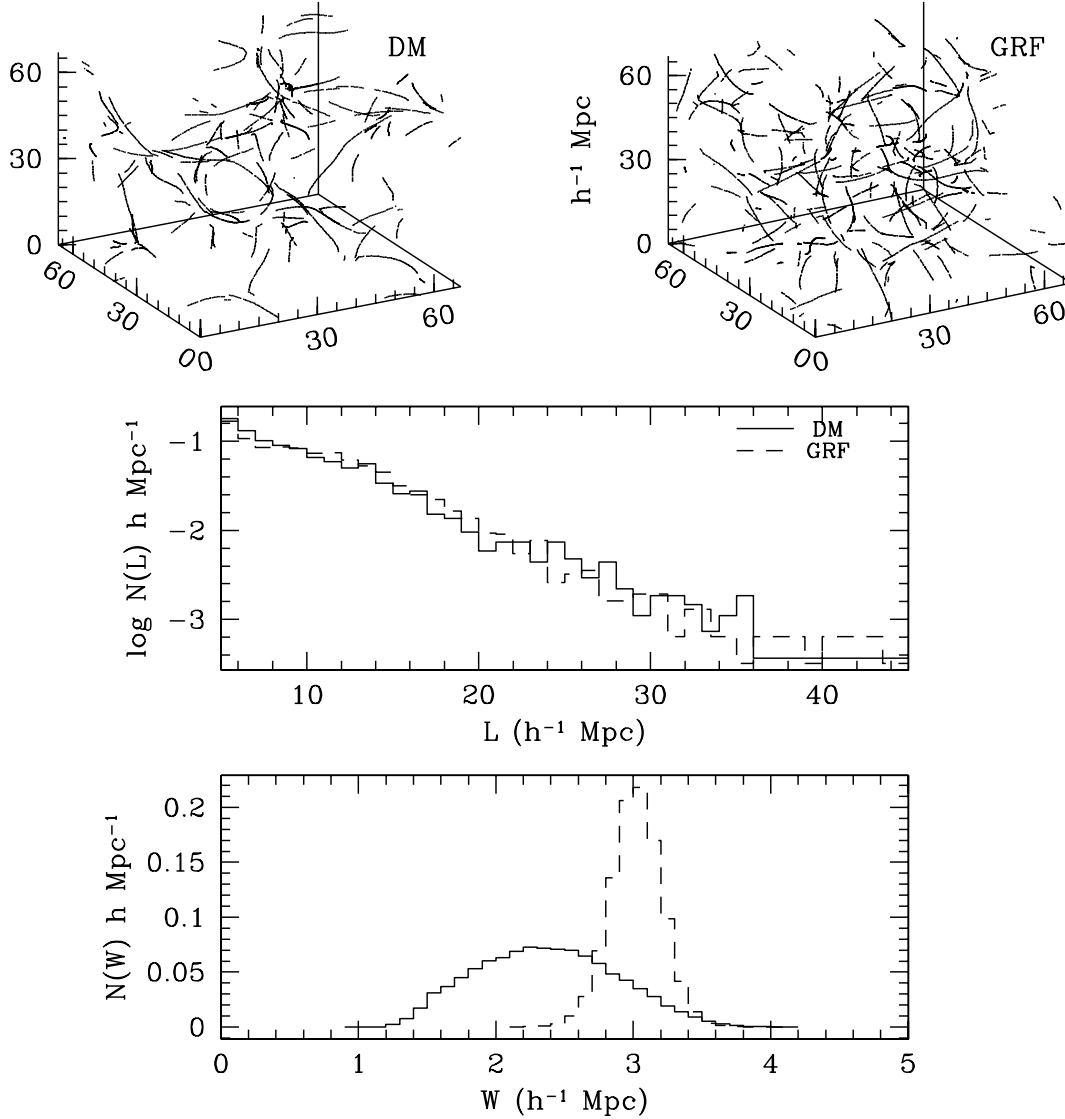
of structure and trace out individual large-scale structures in each distribution.

Using the  $z = 0$  linear power spectrum (Spergel et al. 2007) with corrections in the non-linear regime (Smith et al. 2003; the same one used in the  $N$ -body simulations discussed here), we generate a continuous realization of a three-dimensional Gaussian random field. We use  $l = 5 h^{-1} \text{ Mpc}$ ,  $C = 30^\circ l^{-1}$ ,  $K = 1$  to derive the filament distribution shown in the upper right panel of Fig. 6 and compare it with identical runs on the  $z = 0$  dark matter distribution in the cosmological simulation (upper left panel). The qualitative differences between the two are substantial. While the output for the dark matter distribution resembles a noded network, with filaments converging and ending at vertices in the network, the 'filaments' in the Gaussian random field appear more randomly oriented and show no apparent correlations with one another. Using  $5 h^{-1} \text{ Mpc}$  smoothing, the filament length distributions for the Gaussian random field and dark matter distribution are shown in the centre panel of Fig. 6. The distributions are very similar and clearly exponential above a length of  $\sim 10 h^{-1} \text{ Mpc}$ , with  $N(L) \sim 10^{-0.1L/L}$ , suggesting that filaments have not collapsed much along their longest axis since their formation but have changed their alignment in relation to nearby structures.

We will define the width of a filament element,  $W$ , to be the root-mean-squared perpendicular offset of particles within a smoothing length; that is,

$$W = \sqrt{\frac{\sum_{i=1}^N |\mathbf{R}_i|^2}{N}}, \quad (7)$$

where  $\mathbf{R}_i$  is defined in equation (5) and the sum is over all of the  $N$  particles within one smoothing length of the filament element. In



**Figure 6.** Distribution of filaments in the dark matter distribution (no sparse sampling, upper left panel) and a Gaussian random field (upper right) with the  $\Lambda$ CDM  $z = 0$  non-linear power spectrum. The filaments are found on  $l = 5 h^{-1}$  Mpc scales and we only plot subsections of the full  $200 h^{-1}$  Mpc boxes. In the centre and lower panels, we show the filament length and width distributions, respectively, for the dark matter distribution (solid line) and Gaussian random field (dashed line). In both cases, the filament-finding algorithm was run with  $C = 30^\circ l^{-1}$  and  $K = 1$ .

the lower panel of Fig. 6, we plot the width distributions for the two fields, again using  $l = 5 h^{-1}$  Mpc. The dark matter width distributions are broader and are peaked at smaller widths, suggesting that the filaments have collapsed significantly along two of their principal axes, despite having a similar length distribution. As one would expect with bottom-up structure formation, the width distribution in the Gaussian random field and dark matter distribution are more discrepant at smaller smoothing scales (other scales not shown).

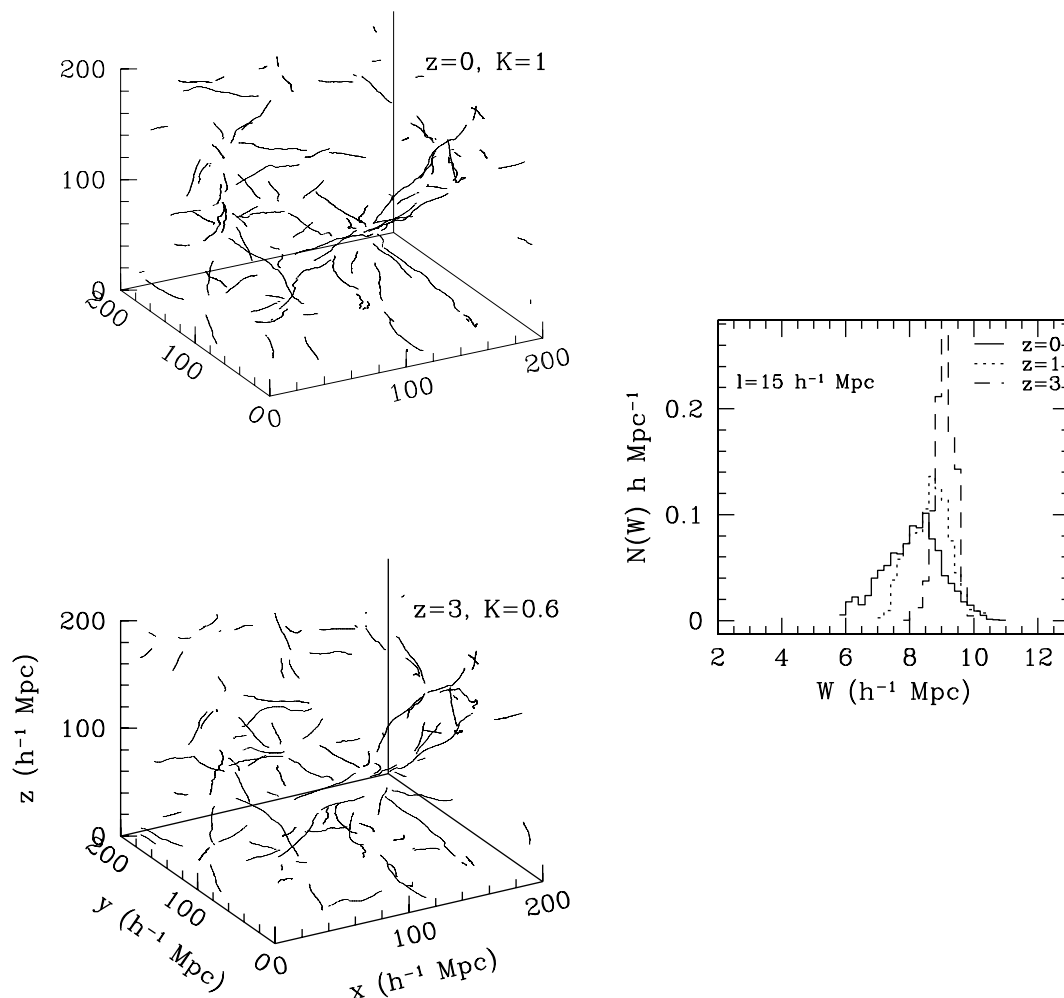
#### 4.1 Filament evolution

In Paper I, we showed that on a given comoving smoothing scale, there was evidence for a wall-to-filament-to-clump evolution with cosmic time. Furthermore, we showed that the axis of structure aligns with the filamentary backbone in two-dimensional slices from cosmological simulations as early as  $z = 3$  (see fig. 14 in Paper I). Fig. 7 shows the filament distribution at  $z = 0$  and  $z = 3$ , now with  $l = 15 h^{-1}$  Mpc so as to test the largest and least-evolved structures

in the simulation box. We used a smaller removal width,  $K = 0.6$ , for the  $z = 3$  filament distribution because the filaments are of lower contrast than at  $z = 0$ , causing equation (3) to overestimate their sizes. The  $z = 3$  and  $z = 0$  filament distributions are very similar to the eye, suggesting that the basic filament framework for  $l = 15 h^{-1}$  Mpc is almost entirely in place at  $z = 3$  [where  $15 h^{-1}$  Mpc fluctuations have  $\langle (\Delta M/M)^2 \rangle^{1/2} \sim 0.1$ ]. The right-hand panel of Fig. 7 shows the filament element width distributions as a function of redshift. As non-linear evolution proceeds, the filament width distributions broaden and peak at smaller widths.

## 5 FILAMENTS IN THE MOCK GALAXY CATALOGUES

Before we proceed to identify filaments in the SDSS data, we run the filament finder on the mock galaxy samples in redshift space (see Paper I) and compare the resulting filaments to those identified in the real-space  $z = 0$  dark matter distribution. The  $l = 5 h^{-1}$  Mpc



**Figure 7.** Dark matter filament distributions at  $z = 3$  and  $z = 0$  after  $l = 15 h^{-1}$  Mpc smoothing. The former was found with a smaller value of  $K$  because the algorithm tends to overestimate filament widths when the filaments are of low contrast (see the text). Both the number and spatial distribution of filaments appear to be similar at the two redshifts, but the width distributions are not, as shown in the right-hand panel. Here, the filament width distributions at  $z = 3$  (dashed line),  $z = 1$  (dotted line) and  $z = 0$  (solid line) are plotted.

filament distribution is very strongly affected by redshift distortions – the contamination rates are typically  $\sim 40$  per cent, about double the contamination of the filament samples without redshift distortions. This is due primarily to the ‘Finger-of-God’ effect, which causes galaxy clusters to extend into narrow, sharp filament-like features along the line of sight. Fortunately, the filament finder is insensitive to these distortions on 10 and 15  $h^{-1}$  Mpc scales because the Fingers-of-God are typically  $\lesssim 1 h^{-1}$  Mpc in width. Nevertheless, we can improve our results if we first remove the Fingers-of-God.

### 5.1 Identification and removal of Fingers-of-God

Fingers-of-God from galaxy clusters are extended along the observer’s line of sight, while real filamentary structure have no preferred direction. In order to separate the Fingers-of-God from the real filaments, we will use an FOF algorithm with two linking lengths:

$$b_{\parallel} = \mathbf{b} \cdot \hat{\mathbf{r}} \\ b_{\perp} = \sqrt{|\mathbf{b}|^2 - b_{\parallel}^2}, \quad (8)$$

where  $\hat{\mathbf{r}}$  is the unit vector along the observer’s line of sight (Huchra & Geller 1982). With these two parameters defined, the algorithm searches for cylindrical structures with a diameter-to-length ratio of  $b_{\perp}/b_{\parallel}$ . Berlind et al. (e.g. 2006, hereafter B06) did an exhaustive study of this two-parameter space and found that  $b_{\perp} = 0.14$  and  $b_{\parallel} = 0.75$  gave unbiased estimates of the group multiplicity function, so we adopt these values in our study.

### 5.2 Filaments after cluster collapse

All of the tests in this section were performed on the samples of mock galaxies with density similar to that of  $M_r < -20$  galaxies. First, we removed galaxy clusters from the *real-space* mock galaxy distribution using an isotropic FOF algorithm with  $b_{\parallel} = b_{\perp} = 0.2$  and a minimum group size of  $N_{\min} = 5$ . For  $l = 10 h^{-1}$  Mpc, the filament incompleteness and contamination rates for the cluster-free filament distribution (33 and 39 per cent, respectively) are much larger than those in real space (16 and 25 per cent), suggesting that overdensities on megaparsec scales are playing an important role in defining filaments on 10  $h^{-1}$  Mpc scales. Similar results are

obtained when clusters are found and removed in redshift space using the approach of Section 5.1.

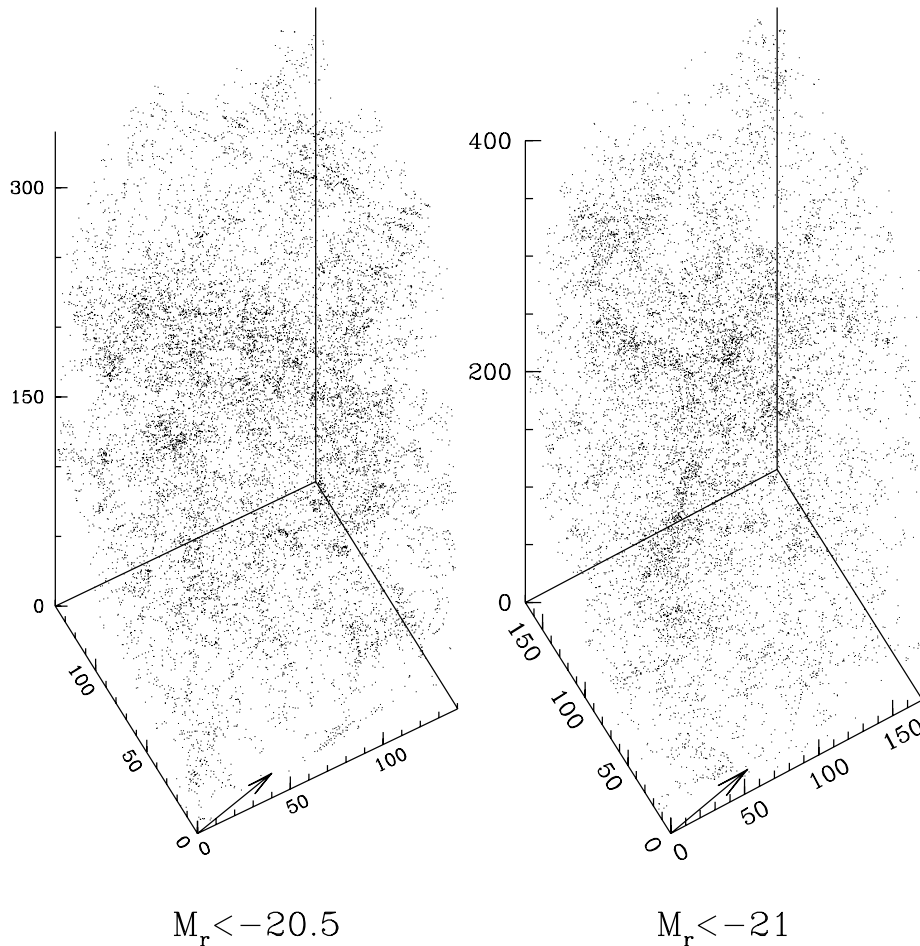
If we instead *collapse* the Fingers-of-God presented by galaxy clusters, we can remove most of the contamination without having to remove the clusters themselves. For this study, we will take the very simple approach of moving all members of a particular cluster to their mean position; that is, we will collapse the Fingers-of-God to a point weighted by the number of galaxies in the cluster. If we follow this procedure, the incompleteness and contamination are smaller (19 and 26 per cent) than the cluster-free mock galaxy distributions and a marginal improvement over the redshift-space distribution with no special treatment of clusters (20 and 27 per cent).

We repeated this exercise for filaments found on a  $5 h^{-1}$  Mpc smoothing scale. Collapsing the Fingers-of-God does lead to a marginal improvement, but filaments are still very poorly defined in redshift space at these densities, with  $\sim 40$  per cent contamination rates. A more sophisticated treatment of the clusters may be needed, but is beyond the scope of this paper. In the section that follows, we will discuss the application of the filament finder to real SDSS data. To minimize contamination, we will be working only with filaments found on 10 and 15  $h^{-1}$  Mpc scales and only after collapsing Fingers-of-God.

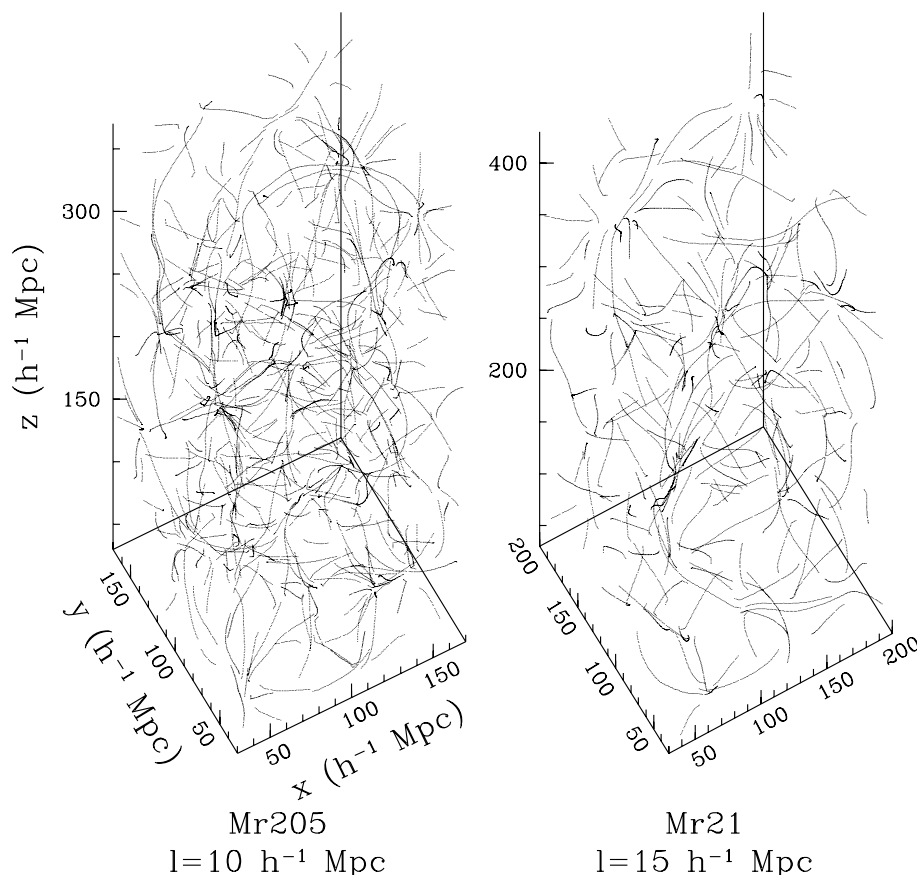
## 6 FILAMENTS IN THE SDSS GALAXY DISTRIBUTION

The SDSS has imaged a quarter of the sky in five wavebands, ranging from 3000 to 10 000 Å, to a depth of  $r \sim 22.5$  (York et al. 2000). As of Data Release 6 (DR6; Adelman-McCarthy et al. 2008), spectra had been taken of  $\sim 800\,000$  galaxies, covering  $9583 \text{ deg}^2$  and extending to Petrosian  $r \sim 17.7$  (Strauss et al. 2002). Galaxy redshifts are typically accurate to  $\sim 30 \text{ km s}^{-1}$ , making it ideal for studies of large-scale structure. For this study, we need a portion of sky with relatively few coverage gaps to minimize the effect of the window function on the  $\lambda$ -space distributions. With this in mind, we construct two volume-limited subsamples from the northern portion ( $8 < \alpha < 16 \text{ h}$  and  $25 < \delta < 60$ ) of the DR6 update of the NYU Value-Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005) the first  $140 \times 140 \times 340 (h^{-1} \text{ Mpc})^3$  in size with  $M_r < -20.5$  ( $Mr205$ ) and the second  $170 \times 170 \times 400 (h^{-1} \text{ Mpc})^3$  in size with  $M_r < -21$  ( $Mr21$ ). The samples extend to maximum redshifts of  $z = 0.12$  and  $z = 0.15$ , respectively, and are plotted in redshift space in Fig. 8. Absolute magnitudes were computed with KCORRECT (Blanton et al. 2003) using SDSS  $r$ -band Petrosian magnitudes shifted to  $z = 0.1$  (and using  $h = 1$ ).

We described the compilation and processing of the SDSS subsamples and their mock counterparts in Paper I. Before generating



**Figure 8.** Two volume-limited samples taken from the SDSS NYU-VAGC large-scale structure sample, with galaxies placed at their comoving positions based on the concordance cosmology. The arrows indicate the location of the Milky Way, which is  $r = (310, -20, 170) h^{-1} \text{ Mpc}$  and  $r = (400, -25, 200) h^{-1} \text{ Mpc}$  in the  $Mr205$  and  $Mr21$  samples, respectively. The  $z$ -axes are parallel to the North Galactic Pole.



**Figure 9.** Filaments found in the *Mr205* (left,  $l = 10 h^{-1} \text{ Mpc}$ ) and *Mr21* (right,  $l = 15 h^{-1} \text{ Mpc}$ ) SDSS samples. These are the full filament samples after the filament finder was run with the ‘best’ parameters. Note that the boxes are different (overlapping) volumes of space and thus are not directly comparable to one another.

the filament distributions, we identify and collapse the Fingers-of-God as described in the last section. After performing this procedure on the *Mr205* and *Mr21* samples (both real and mock), we smooth the former with  $l = 10 h^{-1} \text{ Mpc}$  and the latter with  $l = 15 h^{-1} \text{ Mpc}$ . These choices maximize the volume covered while keeping the sampling rate high enough that filament contamination is under  $\sim 25$  per cent (see Section 5.2).

We run the filament finder on the *Mr205* and *Mr21* galaxy samples using  $C = 40^\circ l^{-1}$  and  $C = 50^\circ l^{-1}$ , respectively, and  $K = 1$ . The resulting filaments are shown in Fig. 9. After removing filaments shorter than a smoothing length, the algorithm finds 489 filaments in *Mr205*, having a total length per unit volume of  $1.9 \times 10^{-3} h^2 \text{ Mpc}^{-2}$  ( $l = 10 h^{-1} \text{ Mpc}$ ), while in *Mr21*, 226 filaments are found with a total length per unit volume of  $7.6 \times 10^{-4} h^2 \text{ Mpc}^{-2}$  ( $l = 15 h^{-1} \text{ Mpc}$ ). For comparison, the mock *Mr205* catalogue contains 451 filaments with a total length per unit volume of  $1.7 \times 10^{-3} h^2 \text{ Mpc}^{-2}$  ( $l = 10 h^{-1} \text{ Mpc}$ ) and the mock *Mr21* catalogue contains 235 filaments with a total length per unit volume of  $8.2 \times 10^{-4} h^2 \text{ Mpc}^{-2}$  ( $l = 15 h^{-1} \text{ Mpc}$ ). Thus, the number density of filaments in the simulations closely matches that in the real Universe.

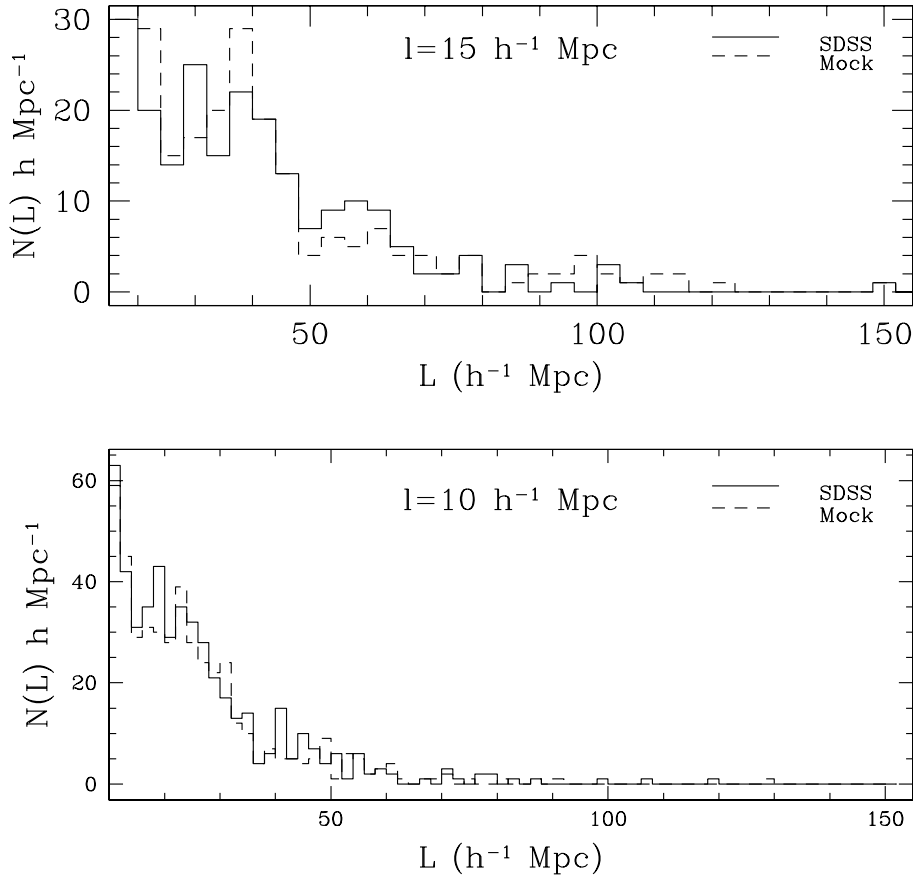
We found in Section 4 that, above two smoothing lengths, dark matter filaments had an exponential length distribution that very closely matched that found in a Gaussian random field with the same power spectrum. This suggests that, even if the filaments in the data are in a different stage of their evolution (i.e. having different  $\sigma_8$ ) than those in the simulations, the length distributions

should be the same between the two. This does appear to be the case, as shown in Fig. 10.

More interesting is the similarity of the width distributions of filament elements, shown in Fig. 11. In the SDSS, we find mean filament widths of 5.5 and  $8.4 h^{-1} \text{ Mpc}$  on 10 and  $15 h^{-1} \text{ Mpc}$  smoothing scales, with standard deviations of 1.1 and  $1.4 h^{-1} \text{ Mpc}$ , respectively. As is demonstrated in Fig. 7, filament element width distributions broaden and shift to smaller widths as non-linear evolution proceeds. A large discrepancy in, for example,  $\sigma_8$  between the simulations and real data should produce filament populations that are at different stages of non-linear evolution and have different width distributions. As such, Fig. 11 suggests that the SDSS filaments are consistent with both the standard model and the set of cosmological parameters used in the simulation.

## 7 RESULTS AND DISCUSSION

This paper develops and uses an algorithm called SHMAFF to identify individual filaments in large-scale structure. In short, it uses the local eigenvectors of the density second-derivative field to define the filament axis and trace individual filaments. Filament ends are defined as points at which the rate of change of the axis of structure exceeds a specified threshold (see Section 2). In a  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmological simulation, this definition produces filament samples that are consistent with our visual impression of structure on a particular scale, are complete with few duplicate detections (Section 3.2) and are robust to sparse sampling (Section 3.4).



**Figure 10.** Length distributions of the  $Mr205$  (lower) and  $Mr21$  (upper) filament samples, with SDSS filaments plotted with a solid line and redshift-space mock catalogues with a dashed line.

In addition to the smoothing scale, the filament finder takes the input parameters  $C$ , the maximum angular rate of change of the filament axis, and  $K$ , the width of filament removal in units of the smoothing length. Using Gaussian smoothing, the ‘best’ values of these input parameters are  $C = 30, 40$  and  $50^\circ l^{-1}$  on  $5, 10$  and  $15 h^{-1}$  Mpc smoothing scales, respectively, and  $K = 1$  for all smoothing scales. After we collapse the Fingers-of-God, contamination and completeness in filament samples found in the mock  $M_r < -20.5$  galaxy distribution are  $\sim 26$  and  $\sim 81$  per cent, respectively. Galaxy clusters are important for defining large-scale filaments and should not be removed before running the filament finder. In redshift space and on smoothing scales above  $\sim 10 h^{-1}$  Mpc, collapsing Fingers-of-God to their mean position produces mock filament samples comparable to those in real space.

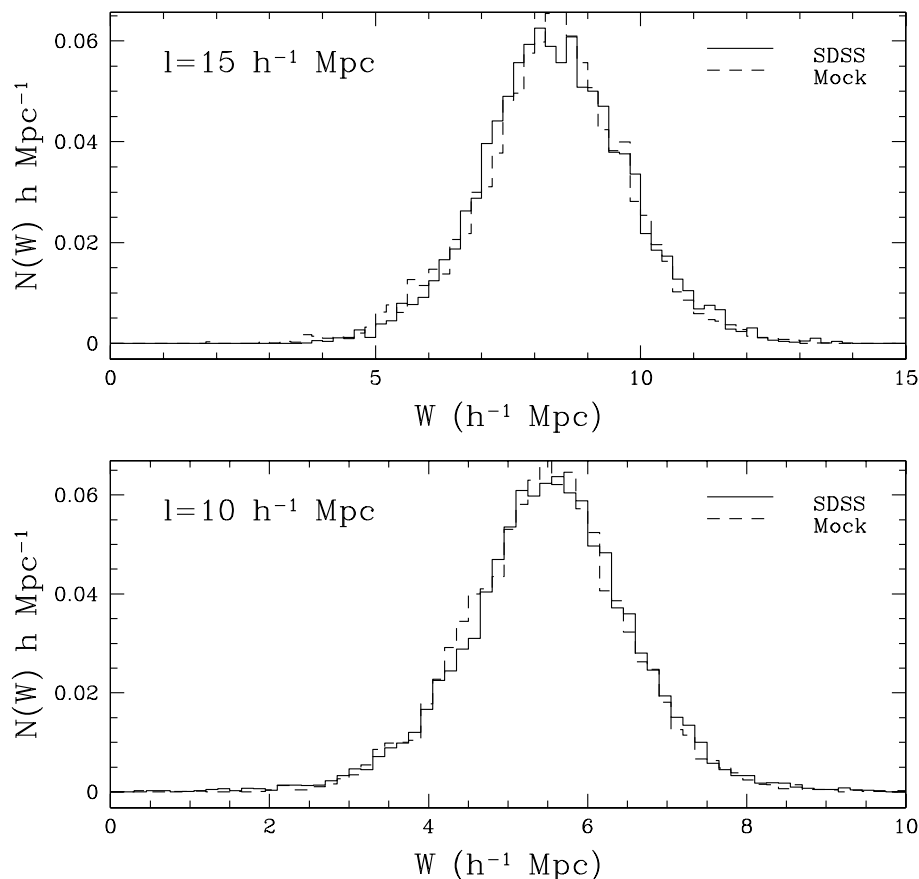
In this paper, we presented two volume-limited subsamples from the northern portion of the SDSS spectroscopic survey (using the NYU-VAGC Catalogue) and computed their filament distributions on  $10$  and  $15 h^{-1}$  Mpc smoothing scales. These distributions were then directly compared to those found in a series of redshift-space mock galaxy catalogues generated from a cosmological simulation using the concordance cosmology. The filament length distributions found in SDSS data are very similar to those found in mock catalogues and are consistent with being drawn from an underlying exponential distribution. The width distributions of filament elements are also very similar between the SDSS data and mock catalogues, suggesting that real filaments are consistent with those in a  $\Lambda$ CDM universe having  $\sigma_8 = 0.85$ ,  $\Omega_\Lambda = 0.71$ ,  $\Omega_m = 0.29$  and  $h = 0.69$ .

Tests on a range of cosmological simulations are needed before this can be turned into a cosmological constraint.

We also generated filament distributions at six redshifts in the output of a  $\Lambda$ CDM cosmological  $N$ -body simulation, from  $z = 3$  to  $z = 0$ . The orientation of the filament network is stable out to  $z = 3$  on comoving smoothing scales at least as large as  $15 h^{-1}$  Mpc. Most of the filaments detected on  $15 h^{-1}$  Mpc scales at  $z = 0$  can be detected at  $z = 3$ . In addition, on a given comoving smoothing scale, filament width distributions shift to smaller widths as the filaments continue to collapse. Narrower filaments will collapse more rapidly, so this also leads to a broadening of the width distributions.

We have demonstrated that our filament finder is able to locate and follow real structures, perhaps most strikingly in Section 4.1, in which we showed that many of the same structures could be seen in a cosmological simulation at both  $z = 3$  and  $z = 0$ . There is some subjective freedom in deciding what constitutes the end of a filament, as no single physical threshold stands out as a discriminator. Nevertheless, we demonstrated in Section 3.2 that the total length of the cosmic network is insensitive to the choice of  $C$  above a certain scale-dependent threshold (once double detections are removed). The minimum value of  $C$  needed to probe the entire filament network may be telling us about the *intrinsic* clumpiness of filamentary structure and may therefore be able to distinguish models of warm dark matter and CDM.

In this paper, we fully developed the SHMAFF algorithm and applied it to the low-redshift galaxy distribution, but there is much that can still be learned from its application to redshift surveys. The filament



**Figure 11.** Width distributions of the *Mr205* (lower) and *Mr21* (upper) filament samples, with SDSS filaments plotted with a solid line and redshift-space mock catalogues with a dashed line. The SDSS filaments in both samples are consistent with those in the cosmological simulations, suggesting that they are in similar stages of non-linear evolution.

evolution seen in cosmological simulations (see Section 4.1) can be tested in the DEEP2 galaxy survey (Davis et al. 2003) at  $z \sim 1$ , and the results of this comparison have already been presented in Choi, Bond & Strauss (2010). On  $l = 5$  and  $10 h^{-1}$  Mpc scales, they confirm a shift in the filament width distribution to smaller widths from  $z \sim 0.8$  to  $0.1$ , as well as a broadening of the filament width distribution. A possible extension of this work is a careful test of the  $\Lambda$ CDM cosmological model, including precision constraints on cosmological parameters, such as  $\sigma_8$ , and tests for primordial non-Gaussianity using the length distribution of filamentary structures. In addition, it would be useful to elaborate on the relationship of large-scale filaments to galaxy clusters and to explore the properties of galaxies in filaments relative to the general galaxy population. Finally, it would be interesting to conduct a careful search for walls in SDSS. Paper I hinted at their presence in the data, but they were only present at low contrast and the  $\lambda$ -space distributions were not optimal for identifying individual wall-like structures.

## ACKNOWLEDGMENTS

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

## REFERENCES

- Adelman-McCarthy J. K. et al., 2008, *ApJS*, 175, 297
- Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, *A&A*, 474, 315
- Aragón-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2008, preprint (arXiv:0809.5104)
- Berlind A. A. et al., 2006, *ApJS*, 167, 1 (B06)
- Blanton M. R. et al., 2003, *AJ*, 125, 2348
- Blanton M. R. et al., 2005, *AJ*, 129, 2562
- Bond J. R., Kofman L., Pogosyan D., 1996, *Nat*, 380, 603
- Bond N. A., Strauss M. A., Cen R., 2010, *MNRAS*, 406, 1609 (Paper I)

- Cen R., Ostriker J. P., 1999, *ApJ*, 514, 1
- Choi E., Bond N. A., Strauss M. A., 2010, *MNRAS*, 406, 320
- Colberg J. M., Krughoff K. S., Connolly A. J., 2005, *MNRAS*, 359, 272
- Colless M. et al., 2001, *MNRAS*, 328, 1039
- Colombi S., Pogosyan D., Souradeep T., 2000, *Phys. Rev. Lett.*, 85, 5515
- Dave R., Hellinger D., Primack J., Nolthenius R., Klypin A., 1997, *MNRAS*, 284, 607
- Davis M., Huchra J., Latham D. W., Tonry J., 1982, *ApJ*, 253, 423
- Davis M. et al., 2003, in Guhathakurta P., ed., *Proc. SPIE Vol. 4834, Science Objectives and Early Results of the DEEP2 Redshift Survey*. SPIE, Bellingham, p. 161
- de Lapparent V., Geller M. J., Huchra J. P., 1986, *ApJ*, 302, L1
- Dietrich J. P., Schneider P., Clowe D., Romano-Díaz E., Kerp J., 2005, *A&A*, 440, 453
- Dolag K., Meneghetti M., Moscardini L., Rasia E., Bonaldi A., 2006, *MNRAS*, 370, 656
- Eisenstein D. J., Hut P., 1998, *ApJ*, 498, 137
- Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, *MNRAS*, 396, 1815
- Gonzalez R. E., Padilla N. E., 2010, *MNRAS*, 407, 1449
- Gott J. R. I., Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, *ApJ*, 624, 463
- Hahn O., Porciani C., Carollo C. M., Dekel A., 2007a, *MNRAS*, 375, 489
- Hahn O., Carollo C. M., Porciani C., Dekel A., 2007b, *MNRAS*, 381, 41
- Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, *MNRAS*, 398, 1742
- Huchra J. P., Geller M. J., 1982, *ApJ*, 257, 423
- Kaiser N., Wilson G., Luppino G., Kofman L., Gioia I., Metzger M., Dahle H., 1998, preprint (astro-ph/9809268)
- Massey R. et al., 2007, *Nat*, 445, 286
- Moody J. E., Turner E. L., Gott J. R. I., 1983, *ApJ*, 273, 16
- Novikov D., Colombi S., Doré O., 2006, *MNRAS*, 366, 1201
- Pimbblet K. A., 2005, *MNRAS*, 358, 256
- Platen E., van de Weygaert R., Jones B. J. T., 2007, *MNRAS*, 380, 551
- Press W. H., Flannery B. P., Teukolsky S. A., 1986, *Numerical Recipes: The Art of Scientific Computing*. Cambridge Univ. Press, Cambridge
- Sahni V., Sathyaprakash B. S., Shandarin S. F., 1998, *ApJ*, 495, L5
- Sathyaprakash B. S., Sahni V., Shandarin S. F., 1996, *ApJ*, 462, L5
- Sathyaprakash B. S., Sahni V., Shandarin S., Fisher K. B., 1998, *ApJ*, 507, L109
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Seldner M., Siebers B., Groth E. J., Peebles P. J. E., 1977, *AJ*, 82, 249
- Sheth J. V., Sahni V., Shandarin S. F., Sathyaprakash B. S., 2003, *MNRAS*, 343, 22
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008a, *ApJ*, 672, L1
- Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008b, *MNRAS*, 383, 1655
- Spergel D. N. et al., 2007, *ApJS*, 170, 377
- Strauss M. A. et al., 2002, *AJ*, 124, 1810
- Tanaka M., Hoshi T., Kodama T., Kashikawa N., 2007, *MNRAS*, 379, 1546
- Thompson L. A., Gregory S. A., 1978, *ApJ*, 220, 809
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.